METHOD FOR CLASSIFYING A DOCUMENT TO BE ASSOCIATED WITH A SERVICE, AND ASSOCIATED SCANNER

METHODS FOR CLASSIFYING A DOCUMENT TO BE ASSOCIATED WITH A SERVICE, AND ASSOCIATED SCANNER

Inventor: Stéphane Manac'h, Rueil Malmaison (FR)

Publication Classification

Int. Cl. H04N 1/46 (2006.01) H04N 1/40 (2006.01)

ABSTRACT

The invention relates to a method for classifying a document (3) to be associated with at least one service (Si), including a step in which a scanner (1) comprising a processor (6) scans (E1) the document (3), wherein said method being characterized in that the method also includes steps in which the processor (6) develops (E2) at least one structure (Σj) representing the document (3), determines (E3) for each service (Si) at least one similitude value (rij) between the structure (Σj) representing the document and a reference structure (Rij) of the same kind and representing said service (Si), deduces (E4) from said similitude value (rij) the service (Si) with which the document (3) is to be associated, and processes (E5) the document (3) according to the service (Si) thus associated. The invention also relates to a scanner for implementing the method.
METHOD FOR CLASSIFYING A DOCUMENT TO BE ASSOCIATED WITH A SERVICE, AND ASSOCIATED SCANNER

GENERAL TECHNICAL FIELD

[0001] The invention relates to a method for classifying a document to be associated with at least one service, comprising a step according to which a scanner including a processor scans the document.

[0002] The invention also relates to a scanner applying said method.

STATE OF THE ART

[0003] Complexification of economical and social activities has caused a spectacular increase in the number of stored and exchanged documents among various economic and social actors, whether these be companies, associations or individuals.

[0004] Thus, it is estimated that a company receives between 50 and 150 different kinds of documents (invoices, complaints . . . ), which in fine involves the processing of thousands of documents per year.

[0005] Most often, these documents are documents either received or sent in the form of paper.

[0006] In order to facilitate exchanges and processing of these documents, it is known how to proceed with scanning and digitizing of these documents in order to convert them into computer files and how to proceed with processing of said files in order to extract relevant information from them.

[0007] This is then referred to as dematerialization of documents.

[0008] Dematerialization may also be used in order to electronically handle data or documents which pass in transit within companies and/or within the scope of exchanges with partners (administrations, clients, suppliers, . . . ).

[0009] Most often, depending on the kind of document, it is necessary to carry out suitable processing for the document. For example, if these are invoices to be processed, the company may resort to a service operator, the role of which is to store and handle these invoices. It will resort to another operator for another type of document.

[0010] In practice, it is then necessary that an employee select a suitable dematerialization software package for this invoice and for the service operator, in order to extract the relevant information for said operator (number of the invoice, identification of the client, amounts, . . . ).

[0011] In this type of method, it is therefore necessary to train employees for processing the documents.

[0012] Further, it is necessary to install and configure powerful software packages within companies for processing these documents.

[0013] All of this increases the processing time and the cost of the processing operation.

[0014] Further, if the service operator desires different processing of the documents, he/she is forced to train the employees again and to apply new on-site parameterization of the dematerialization software packages.

[0015] This type of processing is therefore not very flexible and not very adapted to changes.

[0016] A method for classifying scanned documents, is known from Remy Mullot “Les documents écrits” (written documents), including a semi-supervised learning step.

[0017] Also, a method for classifying scanned documents, including a learning step involving a human operator on the initiative of a machine carrying out the classification is known from George Nagy et al. “Adaptive and interactive approaches to document analysis”.

[0018] The methods of the prior art generally consist of accomplishing preliminary learning, which involves a high commissioning cost, and low robustness to the time-dependent change in documents to be processed.

[0019] Continuous learning as described by Georges Nagy consists of achieving continuous identification of the significant patterns present in the documents to be recognized so as to improve the knowledge base. This assumes that this learning is achieved by an expert. This does not allow it to be achieved at the application level by the non-expert human user and is therefore not applicable to products intended for the market of the company during their current use.

[0020] In every case, the methods of the prior art consist of accomplishing preliminary learning, which involves a high commissioning cost and robustness to the time-dependent change in documents to be processed, which is low.

[0021] The methods of the prior art do not either fully benefit from the powerful analysis capabilities of automatic operators which have to process the documents.

PRESENTATION OF THE INVENTION

[0022] It is therefore necessary to propose an improvement in the processing of documents for overcoming these drawbacks.

[0023] For this purpose, a method according to claim 1 is proposed according to the invention.

[0024] The invention is advantageously completed by the characteristics of the dependent method claims, taken alone or in any of their technically possible combinations.

[0025] The invention also relates to a scanner applying said method.

[0026] The invention has many advantages.

[0027] One advantage of the invention is to propose an economical and flexible solution, the method and device for processing documents may be deployed and transparently updated for a user.

[0028] One other advantage of the invention is to propose an automatic and fast solution for processing documents.

[0029] One other further advantage of the invention is to propose a reliable solution for classifying documents.

[0030] The technique according to the invention, by means of the automatic learning step taking an instruction for either validating the classification by the scanner or not, gives the possibility of fully benefiting from the powerful analysis capabilities of automatic operators which have to process the documents, and which considerably reduce the risks of an error in the classification.

PRESENTATION OF THE FIGURES

[0031] Other features, objects and advantages of the invention will be apparent from the following description, which is purely illustrative and non-limiting and which has to be read with reference to the appended drawings wherein:

[0032] FIG. 1 schematically illustrates a scanner 1 according to the invention;
FIG. 2 schematically illustrates the main steps of a method according to the invention; and
FIGS. 3 illustrate more detailed views of certain steps of FIG. 2.
In the whole of the figures, similar elements bear identical numerical references.

DETAILED DESCRIPTION

FIGS. 2 and 3 schematically illustrate the main steps of a method according to the invention and the main components of a scanner 1 according to the invention.

General Principle

A method for processing a document 3 to be associated with at least one service Si, conventionally comprises a step according to which a scanner 1 scans document 3, during a step E1.
The scanning of the document 3 by the scanner 1 is known to one skilled in the art and is not described in more detail in the following of the present description.
Also, the scanner 1 includes a processor 6 including all the processing and memory storage means known per se for carrying out the steps of the method according to the invention.
The scanning step E1 allows digitization of document 3 in order to convert it into a computer file which may be processed.
Each service Si may for example be a service S1 for processing invoices, a service S2 for processing CVs, a service S3 for processing letters of complaints, a service S4 for processing order forms, etc.
The method then comprises a step E2 during which the processor 6 elaborates at least one structure $\Sigma_j$ representative of document 3.
As this will be seen in more detail in an example in the following of the present description, the structure $\Sigma_j$ representative of document 3 may for example correspond to a list recovering a number of occurrences of at least one word of the document, and/or to a graphical or morphic structure of the document, and/or to a semantic structure of at least one expression of the document. Other structures may also be envisioned.
The method then comprises a step E3 during which the processor 6 determines for each service Si at least one similarity value oij between the representative structure $\Sigma_j$ of document 3 and a reference structure $Rij$ of the same nature and representative of said service Si.
For this purpose, and as this will be seen in more detail in the following of the present description, the scanner 1 includes a memory 5 in which each reference structure $Rij$ and the parameters for the calculation of each value oij are stored.
In FIG. 1, the memory 5 is local and comprised in the scanner 1, but it is understood that the memory 5 may also be remote, for example on a telecommunications network and accessible to the processor 6 by means of conventional communications means 4.
The method then comprises a step E4 during which the processor 6 infers from the values oij the service Si with which document 3 should be associated.
The method then comprises a step E5 during which the processor 6 processes the document 3 according to the thereby associated service Si. The processing operation comprises making the document available for the distant and automatic service operator.

Exemplary Application

A non-limiting exemplary application of the invention is given here.
After the step E1 for scanning the document 3, the processor 6 passes to step E2 during which the processor 6 develops at least one structure $\Sigma_j$ representative of the document 3. For example:
$\Sigma_1$ is a list recovering the number of occurrences of each word of the document 3;
$\Sigma_2$ is a morphic structure of the document recovering at least one logotype optionally present in document 3;
$\Sigma_3$ is a graphical structure of the document recovering at least one graphic organization of the document 3;
$\Sigma_4$ is a semantic structure of the document 3; and $\Sigma_5$ is a structure of the colors of the document 3.
The document may for example be of the typed/printed or handwritten type.
In the case of the typed/printed type of document, character recognition methods (or Optical Character Recognition (OCR)) are well known to one skilled in the art and will not be taken up again in detail in the following of the present description.
The case of the document of the handwritten type, it is also possible now to carry out word recognition as shown for example by the publication of the thesis “Reconnaissance d’écriture manuscrite par des techniques markovienes: une approche bidimensionnelle et generique” (Recognition of handwriting by Markov techniques: a two dimensional and generic approach) of Sylvain Chevalier presented on Dec. 3, 2004. This is then referred to as Intelligent Character Recognition or smart character recognition.

Logotype recognition does not either pose any difficulty as shown in the publication of 1997 “Logo Recognition by Recursive Neural Networks” of E. Francesconi, P. Frasconi, M. Gori, S. Marina, J. Q. Sheng, G. Soda and A. Sperduti, from the universities of Florence, Sienna and Pisa in Italy (http://www.dis.unifi.it/~paciolo/ps/GRFC-97-logo.pdf).
Finally the recognition of a semantic structure or of the colors is also well known to one skilled in the art, as shown for example by the publications.

UCREL Semantic Analysis System (USAS) (http://ucrel.lancs.ac.uk/usas/), and
An example of graphical structure recognition does not either pose any difficulty, as shown by the publication “Analyse et reconnaissance d’images de documents” (Analysis and recognition of images from documents) of Rolf Ingold, published on Aug. 10, 2002 in the “Techniques de l’ingénieur” (reference I7020), or “Reconnaissance de l’imprimé” (Printed Form Recognition) of Philippe Lefevere, published on May 10, 1999 in the “Techniques de l’ingénieur” under reference H1348.
The processor 6 then passes to step E3 during which the processor 6 determines for each service Si at least one similarity value oij between the representative structure $\Sigma_j$ of document 3 and a reference structure $Rij$ of the same nature and representative of said service Si.
For this purpose, the memory 5 includes a database 7 in which are stored reference structures Rij, and for example notably:

R11 is a list of words specific to an invoice, such as for example “Euros”, “amount”, “taxes excluded”, “VAT”, “client”, etc.;

R12 is a morphic structure taking up again at least one particular logo of certain invoices, for example a letterhead of a preferential supplier;

R21 is a list of words specific to a CV, such as for example “school”, “training period”, “experience”, “education”, “job”, etc.;

R23 is a graphic structure taking up again a specific organization of a CV, such as a table with several lines and columns;

R34 is a semantic structure specific to a complaint letter such as for example the semantic connection of the words “object”, “complaint”, “product”, “date”, “purchase”: and

R45 is a structure of the specific colors of an order form, for example from a preferential customer.

The processor 6 determines a similitude value α11, α12, α21, α23, α34, and α45, respectively between the structures 21, 22, 23, 34 and 45 on the one hand and the structures R11, R12, R21, R23, R34 and R45 on the other hand.

For a structure forming a list of words k, the processor for example uses the formula:

$$\sigma_{ij} = \sum_{k \in \Sigma_j} \lambda_k \cdot \delta_k^j(k)(\text{EQ}1)$$

wherein \(\delta_k^j(k)\) has the value

0 if the word k of the list Rij is not in the list \(\Sigma_j\), and

1 if the word k of the list Rij is in the list \(\Sigma_j\), and

wherein \(\lambda_k\) is a weighting coefficient with which more or less significance may be given to the word k.

The parameters \(\delta_k^j(k)\) and \(\lambda_k\) are also stored in the database 7. It is understood that other examples for calculating similitude \(\sigma_{ij}\) are possible for lists of words.

It is also understood that for a morphic structure, a graphic structure, a semantic structure or a color structure, the processor 6 may also use for example the formula (EQ1) according to the presence or not of the logo, of the specific organisation, of similar expressions or colors in the structures of the document 3. Also in this case \(\lambda_k\) is a weighting coefficient by which more or less significance may be given to the relevant element k.

It is understood that other examples for calculating similitude \(\sigma_{ij}\) are possible for a morphic structure, a graphic structure, a semantic structure or a color structure.

The processor 6 then passes to step E4 during which the processor 6 infers from these values \(\sigma_{ij}\) the service SI with which document 3 should be associated.

The index I of the service SI to be associated with document 3 is such that:

$$I = \max(\sigma_{ij})$$

Thus, in our example, if one has

\(\sigma_{21} < \sigma_{11}\)

the list 21 comprises more words in common with list R11 than with list R21, and document 3 is therefore closer to an invoice than to a CV. The associated service is therefore the service SI for processing invoices.

The processor 6 then passes to step E5 during which the document 3 is processed depending on the thereby associated service SI. The processing E5 to be carried out on document 3 and how the data of document 3 are encoded, or even encrypted, actually depends on the operator. The processing operation first of all consists of giving access to the document to an automatic and remote service operator carrying out processing of the document.

Access E5 to the document for the operator comprises, according to a first application, the sending E51 of the document 3 to the operator providing the service SI.

The sending E51 is conventionally carried out by communications means 4 of the scanner 1, known per se by one skilled in the art. They are therefore not described in detail in the following of the present description.

The service operator then carries out, in its location and on proprietary means, the required processing on the document 3 in order to produce the service corresponding to its job. These proprietary means may optionally be very powerful computer means.

The service provided for example by an invoice dematerialization operator is to extract the name of the supplier, references thereof, the amount of the invoice, the date of the invoice, the deadline for payment, the detail of the pieces of information making up the different items of the invoice (unit price, amounts).

The service operator may further for example make sure that the extracted supplier reference corresponds to an existing supplier reference for the relevant client, it may be ensured that the extracted amounts correspond to amounts which are usually processed. This information and its validity are as much elements which help the operator to establish the relevance of the performed dematerialization.

This relevance of the performed dematerialization is a second level, a so-called “application level” for checking the suggested classification.

In order to have a more complete view of the services provided by an operator, it is possible to refer to publications such as the accounting journal, RF Comptable No. 319 of July 2005, or to service tenders proposed by companies such as SERES (dematerialization of incoming invoices, notably).

According to a second possible application of the invention, the access E5 comprises the execution E53 of processing instructions stored in the memory 5 of the scanner 1, so that each operator may integrate the operator’s own specificities into the scanner 1.

For this purpose, the memory 5 includes a memory space 9 dedicated to the storage of processing instructions for the associated service SI. The processing required on the document for producing the service is therefore locally carried out on the scanner 1.

The space 9 is advantageously compartmentalized so that each operator may have the operator’s own memory space and keep control of the operator’s processing.

According to an advantageous alternative of this second application, the processor 6 uploads during a step E52 the processing instructions to the memory 5 of the scanner 1, and more specifically to the space 9, before the execution E53.
of said instructions. This allows the necessary reduction in the memory space, and gives the possibility of making sure that the processor always uses the instructions, directly updated by the operators.

Thus, in a document of several pages for example, the association of the service is achieved on the first page which is locally stored in the memory and processed, the suitable processing instructions may then be uploaded in hidden time during the scanning of the second page for example.

The upload conventionally takes place by using communications means.

Advantageously, the inference step comprises learning, referenced by a figure, in which the processor proposes, during a step for the service to be associated with the first page, the basis of calculated similitudes.

For this purpose, the scanner includes display means conventionally including a liquid crystal screen, for example allowing display of the proposal of step by the name of the associated service Si.

In our example, it may be used for example, that the processor associates the document with the service.

The learning process further includes a step during which a control sends an instruction for validating the proposal, for example referenced by OK, or an instruction invalidating the proposal, for example referenced by NXT or CNCL.

For this purpose, the control may include an interface, for example a push button and/or a tactile option, associated with the means.

If after a study of document, the human user estimates that the proposal of processor is correct, he then actuates the control for sending a validation instruction of the OK type.

On the contrary, the user estimates that the proposal from the processor is not correct, he/she then actuates the control for sending a non-validation instruction.

The non-validation instruction may be of two types.

The first type is referenced by NXT, and may for example be an instruction requiring that the processor propose the service which has the next highest similitude, and so forth until a correct proposal which generates the sending of a validation instruction, as explained above.

The second type is referenced by CNCL, and may for example be an instruction requiring that the processor cancel the service proposals, the association of the service with the document then being for example carried out manually by the user.

The processor takes into account, during a step, said instruction in subsequent service proposals.

For this purpose, the database includes a correspondence table between the values and the similitudes in order to either retrieve validations or not for each proposal of the service Si.

In the case when the user accepts by an OK instruction, the proposal of the service Si, the processor then carries out a mathematical iteration for example during step so that:

\[ \phi' = f(\phi) \]

wherein a finite value \( \phi \) in the table is replaced with a new finite value \( \phi' \) by means of \( f \), a function such that:

\[ f(x) = x^2 \]

The function \( f \) may for example be an incremental function such as:

\[ f(x) = x + 1 \]

In the case when the user does not accept the proposal of the service Si by a NXT instruction, the processor does not modify the correspondence table.

It is then understood that the processor may advantageously promote in subsequent proposals, services having previously led to the sending of an OK validation instruction by the control.

For this purpose, equation (EQ1) is modified into (EQ2) such that:

\[ \sigma_{ij} = \Phi_i \cdot \sum_{k \in \mathbb{R}} \lambda_k \cdot \delta_{ij}(k) \]

The instruction of validation or not may, also advantageously as an addition to the instructions sent by the human operator, be sent by the automatic and remote service operator, once it has access to document. The operator is actually often equipped with powerful analysis equipment, and may better detect whether the document just received corresponds or does not correspond to the service thereof. The operator may therefore warn the scanner on the quality of the proposal E41. The control is in this case associated with communications means.

An example of the main steps carried out by the processor and notably the steps of a learning process carried out by the processor are developed hereafter.

---

# The learning process uses an input parameter,
# i.e. a text file containing the result of a character recognition
# processing operation OCR on the original document.
# The following imported functions (pickle and sys) are standard
# # functions of the programming language 'Python', and they are
# # therefore known to one skilled in the art.
# import pickle
# import sys
# Extract of the list of words contained in the document
structure = set(open(sys.argv[1], "r").read().lower().split())
# Loading the knowledge base, stored in base 7.
try:
services = pickle.load(file("services.db", "rb"))
except:
    # In the case of an error, reset with default values
    services = [
        "Invoice": ["Invoice": 90,
        "Statement": 5,
        "Var": 100,
        "cash-free": 80,
        "Money": 10,
        "total": 90,
        50]
    "letter": ["clear": 80,
        "Sir": 50,
        "Madam": 50,
        "Regards": 33,
        "Sincerely": 33,
        "Faithfully": 33,
        50]
    "Slip": ["Delivery": 75,
        "Date": 10,
        "Value": 80,
        "Supplier": 50,
        "Reference": 70,
        "heading": 60,
        50]
#Calculate the score of each service
scores = []
for name, (ref, phi) in services.items():
    score = 0
    for word, lamb in ref.items():
        if word in structure:
            score += lamb * scores.append((score * phi, name))
#SEARCH AND LEARNING PROCESS: Browse through the list in
#decreasing score order until the user confirms
ok = False
for service in reversed(sorted(scores)):
    print "Check that the document relates to service \%s\% service"
    response = raw_input("Enter O to confirm, N to reject:
        ").lower()
    if response == "o":
        print "You have confirmed that the document relates to
the service \%s\% service"
        # Update phi for the current service
        services[service] += 1
        # New possible version:
        services[service] = (services[service] + 100) // 2
        ok = True
    else:
        # Update phi for the current service
        services[service] -= 1
        # Other possible version:
        services[service] = (services[service] + 1) // 2
        if not ok:
            print "Document does not relate to any known service!"
            # SAVE LEARNING PROCESS: Save the knowledge base
            in base 7
            print "Values of Phi": [\"%s\%\%d\% service, phi\) for (service,
                (_, phi)) in services.items()]
            pickle.dump(services, file("services.db", "wb"))
            raw_input("Press Enter to continue")

[0120] For example, in the case of a service operator ensuring dematerialization of invoices, if the automatic extraction of information allows it to find relevant information (identification of an already known supplier, of usual amounts, of a global amount which is actually the sum of the amount of the different items), it may send a validation instruction; on the contrary if such elements are not extracted, it may inform the scanner that the wrong service was proposed.

[0121] The sending of the instruction by the operator may either be combined or not with the sending of the instruction by the human user.

[0122] Of course, any other learning process may also be applied.

[0123] Preferentially, the processor 6 compares the similarity value rij with a confidence threshold t. The processor 6 considers that when the obtained value rij is greater than the threshold t, such as rij > t, the confidence in the recognition of the type of document is sufficient so as not to await a validation instruction or not on the behalf of the user or of the operator. The learning step E40 is therefore interrupted.

1. A method for classifying a document (3) to be associated with at least one service (Si), comprising:
   a) a step according to which a scanner (1) includes a processor (6) scans (E1) document (3), and the steps according to which the processor (6) develops (E2) at least one structure (Σj) representative of the document (3) compares the representative structure (Σj) of the document with at least one reference structure (Rij) of the same nature and representative of said service (Si), determines (E3) for each service (Si) at least one similarity value (rij) between the representative structure (Σj) of the document and the reference structure (Rij) of the same nature and representative of said service (Si), and infers (E4) from this similarity value (rij) the service (Si) with which the document (3) is to be associated, and thereby forms itself a local proposal (E41) for classifying it,

   wherein the inference (E4) comprises a learning process (E40) according to which:
   a) a human user and/or a thereby associated remote and automatic service operator send back (E42) to the processor (6), via a control (8), an instruction for validation or invalidation of the proposal (E41) locally carried out by the processor, and
   b) the processor (6) takes into account said instruction (E43) in subsequent service proposals (E41), and promotes in subsequent proposals (E41) services having led to the sending (E42) of a validation instruction (OK) by the control (8).

2. The method according to claim 1, wherein the processor (6) compares the similarity value (rij) to a confidence threshold (t).

3. The method according to claim 1, wherein the processor gives access (E5) to the document (3) to the thereby associated, distant and automatic service operator (Si) depending on the service (Si).

4. The method according to claim 3, wherein the access (E5) to the document comprises the sending (E51) of the document (3) to the operator providing the service, through communication means (4) of the scanner (1).
5. The method according to claim 3, wherein the access (E5) to the document comprises the execution (E53) of processing instructions stored locally in a memory (5) of the scanner (1).

6. The method according to claim 5, wherein the processor (6) uploads (E52) the processing instructions towards the memory (5) of the scanner.

7. The method according to claim 1, wherein the structure (Xj) representative of document (3) corresponds:
to a list taking up again a number of occurrences of at least one word of the document, and/or
to a graphic or morphic structure of the document, and/or
to a structure of the colors of the document, and/or
to a semantic structure of at least one expression of the document.

8. A scanner (1) for a document (3) and for selecting services to be associated with the document (3) comprising a processor (6), said processor (6) being adapted for:
developing (E2) at least one structure (Xj) representative of document (3),
determining (E3) for each service (Si) at least one similitude value (rij) between the representative structure (Xj) of the document and a reference structure (Rij) of the same nature and representative of said service (Si), and locally inferring (E4) from this similitude value (rij) the service (Si) with which document (3) is to be associated, wherein the processor (6) is adapted for
a learning process (E40) during the inference (E4),
and is thus suitable for receiving, from a human operator and/or from a thereby associated, distant and automatic service operator, an instruction for validating or invalidating the proposal (E41) locally carried out by the processor, via a control (8) and
for taking (E43) said instruction into account in subsequent service proposals (E41) by the processor (6), and is thus adapted for promoting in subsequent proposals (E41) of the services having led to the sending (E42) of a validation instruction (OK) by the control (8).

9. The scanner according to claim 8, further comprising remote communication means (4) with a server and/or with the service operator.

10. The method according to claim 2, wherein the processor gives access (E5) to the document (3) to the thereby associated, distant and automatic service operator (Si) depending on the service (Si).

11. The method according to claim 2, wherein the structure (Xj) representative of document (3) corresponds:
to a list taking up again a number of occurrences of at least one word of the document, and/or
to a graphic or morphic structure of the document, and/or
to a structure of the colors of the document, and/or
to a semantic structure of at least one expression of the document.

12. The method according to claim 3, wherein the structure (Xj) representative of document (3) corresponds:
to a list taking up again a number of occurrences of at least one word of the document, and/or
to a graphic or morphic structure of the document, and/or
to a structure of the colors of the document, and/or
to a semantic structure of at least one expression of the document.

13. The method according to claim 10, wherein the structure (Xj) representative of document (3) corresponds:
to a list taking up again a number of occurrences of at least one word of the document, and/or
to a graphic or morphic structure of the document, and/or
to a structure of the colors of the document, and/or
to a semantic structure of at least one expression of the document.

14. The method according to claim 4, wherein the structure (Xj) representative of document (3) corresponds:
to a list taking up again a number of occurrences of at least one word of the document, and/or
to a graphic or morphic structure of the document, and/or
to a structure of the colors of the document, and/or
to a semantic structure of at least one expression of the document.

15. The method according to claim 5, wherein the structure (Xj) representative of document (3) corresponds:
to a list taking up again a number of occurrences of at least one word of the document, and/or
to a graphic or morphic structure of the document, and/or
to a structure of the colors of the document, and/or
to a semantic structure of at least one expression of the document.

16. The method according to claim 6, wherein the structure (Xj) representative of document (3) corresponds:
to a list taking up again a number of occurrences of at least one word of the document, and/or
to a graphic or morphic structure of the document, and/or
to a structure of the colors of the document, and/or
to a semantic structure of at least one expression of the document.

* * * * *